Low-rank matrix completion and denoising under Poisson noise

Andrew D. McRae and Mark A. Davenport
Georgia Institute of Technology

***Abstract*—This paper considers the problem of estimating a nonnegative low-rank matrix from noisy Poisson observations of all or a subset of its entries. Specifically, we analyze an estimator defined by a constrained nuclear-norm minimization program. We derive a high-probability upper error bound (in the Frobenius norm metric) that depends on the matrix rank, the fraction of entries observed, and maximal row and column sums of the true rate matrix. We furthermore show that this bound is (within a constant) minimax optimal in classes of matrices with low rank and bounded row and column sums.**

I. Introduction

In this paper, we consider the problem of estimating a nonnegative matrix $M \in \mathbf{R}^{m \times n}$ given independent observations distributed according to $\mathrm{Poisson}(M_{ij})$ for $(i,j) \in \Omega$, where Ω is a (not necessarily strict) subset of $\{1, \dots, m\} \times \{1, \dots, n\}$. If we do not make an observation for every entry of a matrix, the recovery problem is, in general, ill-posed if we do not make any assumptions about what kind of matrix we expect to recover. A standard assumption for this type of problem is that the unknown matrix M is *low-rank*; i.e., the dimension of the spans of the columns and rows of M is much smaller than the actual numbers of columns and rows. This assumption greatly reduces the number of degrees of freedom in the model, making the recovery problem more tractable. Even if we do observe every entry, we can exploit the structure of the model to reduce the error due to Poisson noise.

While the problems matrix completion and denoising have received a significant amount of attention in the settings of Gaussian noise and of small, bounded (in ℓ_2) perturbations (e.g., [1]–[3]), the case of Poisson noise has received comparatively less attention. The Poisson noise model is often much more natural in applications where the observations arise via some form of counting process. The ability to recover (or denoise) a low-rank signal from noisy, count-based observations is useful in many situations; we briefly mention two examples.

One widespread application is imaging. This includes conventional cameras (which often suffer from noise in low light or with short exposures), but it also includes 3-D imaging methods such as X-ray computed tomography (CT) and positron emission tomography (PET), which, in medical imaging, would greatly benefit from an improved noise/radiation dose tradeoff. In these scenarios, the Poisson noise model is natural because the observations consists of counts of particle (e.g., photon) arrivals at a detector. In many of these settings, such as when observing a periodic or slowly-varying sequence of images, a low-rank assumption on the underlying data is natural (see, e.g., [4] for an overview of low-rank modeling in image applications).

Another interesting application is topic modeling, which is a common form of dimensionality reduction for text documents. In this case, our observations consist of counts of word occurrences in a corpus of documents. If we suppose that these documents can be decomposed according to a small set of topics, and that within each topic documents will exhibit similar word occurrence counts, then a low-rank assumption on the word-count matrix is natural. For example, the popular PLSI model [5] uses similar assumptions. We note, however, that existing algorithms have no guarantees of performance.

A. Summary of main results

In our analysis, we assume a Bernoulli sampling of the matrix entries: i.e., the events $\{(i,j) \in \Omega\}$ are independent with probability $p \in (0,1]$, and the observed Poisson random variables are independent conditioned on Ω. Note that taking $p = 1$ handles the case in which we observe every entry of the matrix.

Let $\mathcal{A}_\Omega \colon \mathbf{R}^{m \times n} \to \mathbf{R}^\Omega$ denote the entry-wise sampling operator given by $(\mathcal{A}_\Omega(Z))(i,j) = Z_{ij}$ for $(i,j) \in \Omega$. Given Poisson observations $X \sim \mathrm{Poisson}(\mathcal{A}_\Omega(M))$, we consider an estimator of the form

$$\widehat{M}^\delta = \operatorname*{argmin}_{M' \in [0,\infty)^{m \times n}} \|M'\|_* \text{ s.t. } \|\mathcal{A}_\Omega^*(X) - pM'\| \le \delta, \quad (1)$$

where $\delta > 0$ is a parameter which will be chosen so that the true rate matrix M is feasible. Here, for a matrix A, $\|A\|$ and $\|A\|_*$ denote the operator norm and nuclear norm of A, respectively. Note that (1) is a convex optimization problem (a semidefinite program, to be precise), so $\widehat{M}^\delta$ is tractable to compute.

Denoting by $\|A\|_F$ the Frobenius norm of a matrix A, Theorem 1 states that, if M has rank r, and δ is properly chosen, we have, with high probability,

$$\|M - \widehat{M}^\delta\|_F \lesssim \sqrt{\frac{r}{p}}\widetilde{\sigma}(M) + \text{logarithmic terms}, \quad (2)$$

where

$$\widetilde{\sigma}(M) = \max_i \sqrt{\sum_j M_{ij} + (1-p)M_{ij}^2} + \max_j \sqrt{\sum_i M_{ij} + (1-p)M_{ij}^2}.$$

In many situations (see Section II-C), the logarithmic term is negligible, so we can approximate this result by the bound

$$\|M - \widehat{M}^\delta\|_F \lesssim \sqrt{\frac{r}{p}}\widetilde{\sigma}(M). \quad (3)$$

We then use two standard methods to find lower bounds on the minimax risk of any estimator in classes of matrices with bounded row and column sums. These results (Theorems 2 and 3) can be summarized as follows: over all nonnegative matrices M such that $\operatorname{rank}(M) \le r$, and $\widetilde{\sigma}(M) \le \sigma$, we have

$$\inf_{\widehat{M}} \sup_{M} \mathbf{E}_M \|M - \widehat{M}\|_F \gtrsim \sqrt{\frac{r}{p}}\sigma$$

Thus, Theorem 1 is optimal (up to a multiplicative constant and an additive logarithmic factor) for this class of matrices.

To gain a more intuitive understanding of our result, it is helpful to examine the formula for $\widetilde{\sigma}$. For simplicity, assume, without loss of generality, that the row sums dominate the column sums, so that

$$\widetilde{\sigma} \approx \max_i \sqrt{\sum_j M_{ij} + (1-p)M_{ij}^2}.$$

The two terms inside the sum have different roles. The first term (M_{ij}) corresponds to the variance of the Poisson random variables. Indeed, if we take $p = 1$, this is the only term, so our result has the form

$$\|M - \widehat{M}^{\delta}\|_F \lesssim \sqrt{r}\left(\max_i \sqrt{\sum_j M_{ij}}\right).$$

If we do not impose any structure on the model, the maximum likelihood (and least-squares) estimate is $\widehat{M}^{\text{MLE}} = X$, which has risk

$$\mathbf{E}\|M - \widehat{M}^{\text{MLE}}\|_F^2 = \sum_{i,j} M_{ij}.$$

If every row of M has approximately the same sum, the estimate $\widehat{M}^{\delta}$ improves on $\widehat{M}^{\text{MLE}}$ (in squared Frobenius error) by a factor of approximately r/n. If the sums of the rows of M differ significantly, the improvement is smaller. However, this should not be too surprising—if the variance in the problem is already concentrated into a smaller sub-matrix, we are effectively solving a smaller problem, and hence the low-rank assumption is less restrictive and, therefore, less helpful.

The second term (of the form $(1-p)M_{ij}^2$) in the formula for $\widetilde{\sigma}$ corresponds to the inherent difficulty in estimating the values of a matrix due to the fact that we do not observe every entry. This term in the lower bound applies regardless of the noise model, even when there is no noise. This might seem to contradict existing exact noiseless matrix completion results, but we note here that such results assume additional structure (incoherence of the row and column spaces) beyond what we are assuming here. In fact, the matrices used in the proof of Theorem 3 are highly coherent.

Although this second error term is necessary for general matrices, an interesting open problem is whether it could be entirely removed (leaving only the variance term) when we assume additional structure (such as incoherence) on the true rate matrix. Such a result would be a bridge between existing noisy and noiseless matrix completion literature; the existence of exact completion for the noiseless case implies that current results for the noisy case (including this paper) become useless when the signal-to-noise ratio goes to infinity. An exception is [3], but we note that this approach is not without its own drawbacks as this approach leads to error rates which are suboptimal with respect to the rank r.

B. *Comparison to prior work*

There are several categories of existing literature to which we can compare our results. Some papers explicitly consider Poisson noise, using a maximum-likelihood framework. Cao and Xie [6] consider nuclear-norm penalized maximum likelihood for matrices contained in a nuclear norm ball (rather than exactly low-rank matrices). This approach uses an empirical process argument to bound the Kullback-Leibler divergence between the true and predicted distributions. This argument requires a Lipschitz condition on the log-likelihood function, which, for the Poisson distribution, requires imposing a lower bound on the rates. Soni, Jain, Haupt, *et al.* [7] and Soni and Haupt [8] consider a penalized maximum likelihood estimator from a carefully-chosen finite set of candidates (which is exponentially large in the size of the problem and hence computationally intractable). The matrices considered have a nonnegative low-rank factorization (with a particular emphasis on the case when one factor is sparse). They use an information-theoretic argument to bound the expected error in terms of Bhattacharyya distance. The result of [7], which applies to matrix completion, requires imposing a lower bound on the rates, while that of [8], which considers only denoising, does not. All three papers find an upper bound on Frobenius error in terms of the statistical error metrics that they originally bound.

Other, more general approaches, are designed specifically with Frobenius-norm error in mind. One class of methods uses “restricted strong convexity” arguments, introduced by Negahban and Wainwright [1]. These methods rely on approximating the Frobenius norm in certain restricted classes of matrices using only samples of the entries. Another class of methods, which was first introduced by Koltchinskii, Lounici, and Tsybakov [2], served as the inspiration of the approach we take in this paper; the estimator in their paper (which is somewhat more general than ours) is similar to a nuclear-norm regularized least-squares estimate, but with certain random terms replaced by their known expectations. These two classes of methods rely heavily on bounding the operator norms of two different kinds of random matrices that arise from the sampling pattern.

An interesting blend of techniques can be seen in the works of Lafond [9] and Gunasekar, Ravikumar, and Ghosh [10], who combine some of the general approaches mentioned above with maximum likelihood estimation for exponential families of distributions. These methods, like those in [6] and [7], are difficult to apply to the Poisson distribution without imposing a lower bound on rates because, as the mean λ of the distribution goes to 0, the “natural parameter” $\log \lambda$ goes to $-\infty$, whereas the methods in [1] and [2] generally require parameters to be bounded. They also require (approximate) low rank in the matrix of natural parameters; in the Poisson case, this is equivalent to assuming a bound on the rank of the matrix $[\log M_{ij}]$ of elementwise logarithms of the means, which is somewhat non-standard, and certainly not the same as assuming a bound on the rank of the original matrix M.

Most of the papers mentioned above do not find error bounds which explicitly depend on the “true” rate matrix; rather, they find uniform upper bounds for classes of structured matrices with uniform upper (and, sometimes, lower) bounds on the entries. To compare our results directly to this literature, we consider what we obtain when we only impose a uniform upper and lower bounds (by, say $\lambda_{\max}$ and $\lambda_{\min}$) on the matrix entries. The approximate bound of (3) reduces to

$$\|\widehat{M} - M\|_F^2 \lesssim (\lambda_{\max} + (1-p)\lambda_{\max}^2)\frac{rm}{p},$$

where we have assumed, without loss of generality, that $m \geq n$. Previous results show similar error rates in terms of matrix dimensions for exactly low-rank matrices. For example, [7] establishes a bound of

$$\mathbf{E}\|\widehat{M} - M\|_F^2 \lesssim \frac{\lambda_{\max}^3}{\lambda_{\min}}\frac{rm}{p}\log m,$$

which provides a similar dependence on r, m, and p, but with an additional logarithmic term and a worse dependence on the minimum and maximum matrix values. In a slightly different setting, [6] shows that for matrices in the nuclear norm ball of radius $\lambda_{\max}\sqrt{rmn}$ (which is a convex relaxation of the exact low-rank constraint), we instead obtain (ignoring logarithmic terms and a complicated but severe dependence on $\lambda_{\max}$ and $\lambda_{\min}$) an error bound of

$$\|\widehat{M} - M\|_F^2 \lesssim \frac{\sqrt{rn}m}{\sqrt{p}},$$

where p is now the number of samples for entry in a uniform-at-random sampling model. The different dependence on r and p is interesting, but, if one compares it to results in linear regression over ℓ_1 balls (see, e.g., [11]), the rate given is perhaps not surprising.

Because this paper studies minimax error rates in Frobenius-norm error, it uses techniques which do not rely heavily on the Poisson nature of the noise; the methods we present for upper bounding the error would apply to any problem with sub-exponential noise, and we could similarly find matching lower bounds in problems with many different noise distributions. For this specific error metric, this gives us an advantage over more distribution-specific approaches such as [6], [7], [9], [10], in part because we do not have to approximate the Frobenius norm error by a statistical divergence measure or by a norm in a transformed parameter space. Our results also do not suffer from the fact that a Poisson distribution's likelihood function is ill-conditioned for very small rates. In addition, our results avoid a multiplicative logarithmic factor in the error upper bounds (replacing it with an additive factor that is often negligible); this achievement is almost entirely due to the use of recent results in bounding the operator norm of a random matrix (such as [12]).

Finally, much of the previous literature in the Poisson case (from those mentioned above, [6], [7], [9]) finds lower bounds on minimax risk in certain classes of matrices. Although these lower bounds have the same large-scale error rate (in terms of the rank and dimensions of the matrix and the number of samples) as the corresponding upper bounds, they differ from the upper bounds by factors that are logarithmic in the problem size and that depend on the ratio of largest to smallest allowable rates. To our knowledge, the results in this paper are the first for noisy low-rank matrix completion in which the minimax rate for large classes of matrices is found to within a universal constant.

We add a final caveat to our results by noting that $\|\widehat{M} - M\|_F$ might not always be the most appropriate error metric; for example, there is a much larger difference qualitatively (and quantitatively, if we use an appropriate statistical divergence) between Poisson distributions of means 0 and 10 than between Poisson distributions of means 100 and 110. Further investigation of distribution-specific methods (such as maximum likelihood) that yield bounds in more statistically-motivated metrics is thus certainly warranted.

II. Theorem Statements and Proof Sketches

A. Upper bound

In this section, we present and sketch the proof of our main result, which is an upper bound on estimator error. Complete proofs will appear in a future publication.

Theorem 1. *Let M be a non-negative $m \times n$ matrix with rank r. Let $\lambda_{\max} = \max_{ij} M_{ij}$, and let*

$$\begin{aligned}\widetilde{\sigma} = &\max_i \sqrt{\sum_j M_{ij} + (1-p)M_{ij}^2} \\ &+ \max_j \sqrt{\sum_i M_{ij} + (1-p)M_{ij}^2}.\end{aligned}$$

Suppose $\Omega \subset \{1, \ldots, m\} \times \{1, \ldots, n\}$ is chosen according to a Bernoulli sampling model with sampling probability p, and suppose, conditionally on Ω, $X \sim \operatorname{Poisson}(\mathcal{A}_\Omega(M))$. Set $\epsilon \in (0, 1/2)$ and choose δ such that

$$\begin{aligned}\delta \geq &2\sqrt{p}\widetilde{\sigma} + \frac{8\epsilon}{\sqrt{mn}} \\ &+ C\max\left\{\lambda_{\max}, 4\log\frac{2mn}{\epsilon}\right\}\sqrt{\log\frac{m \vee n}{\epsilon}},\end{aligned} \tag{4}$$

where C is a universal constant. Let $\widehat{M}^\delta$ be the solution of (1)*. With probability at least $1 - 2\epsilon$,*

$$\|\widehat{M}^\delta - M_0\|_F \leq \frac{4\sqrt{2r}\delta}{p}. \tag{5}$$

The result follows from a series of lemmas. The first step in upper bounding the error is the following (deterministic) result:

Lemma 1. *Suppose M is a rank-r matrix such that $\|\mathcal{A}_\Omega^*(X) - pM\| \leq \delta$. Then*

$$\|\widehat{M}^\delta - M\|_F \leq \frac{4\sqrt{2r}\delta}{p}. \tag{6}$$

The proof is a standard argument based on the optimality of $\widehat{M}^\delta$ for (1) and the feasibility of M.

To bound the error of our estimator, it remains to find an upper bound on the random operator norm $\|\mathcal{A}_\Omega^*(X) - pM\|$ when $X \sim \operatorname{Poisson}(\mathcal{A}_\Omega(M))$. We will use the following fundamental lemma, which was originally proved by Bandeira and van Handel [12] and appears with a slightly improved constant in [13]:

Lemma 2 (Theorem 4.9 and Remark 4.11 in [13])**.** *Let X be a random $m \times n$ matrix whose entries are independent, centered, and almost surely bounded in absolute value by a constant b. Let*

$$\sigma = \max_i \sqrt{\sum_j \mathbf{E}\, X_{ij}^2} + \max_j \sqrt{\sum_i \mathbf{E}\, X_{ij}^2}.$$

Then

$$\mathbf{P}(\|X\| \geq 2\sigma + t) \leq (m \vee n) \exp\left(-\frac{t^2}{C_1 b^2}\right),$$

where C_1 is a universal constant.

Poisson random variables are clearly unbounded, so Lemma 2 does not directly apply. The following technical lemma (which is proved by considering truncated matrix elements) allows us to extend the result to the case of random variables with sub-exponential tails.

Lemma 3. *Let X be a random $m \times n$ matrix whose entries are independent and centered, and suppose that for some $v, t_0 > 0$, we have, for all $t \geq t_0$,*

$$\mathbf{P}(|X_{ij}| \geq t) \leq 2e^{-t/v}.$$

Let $\epsilon \in (0, 1/2)$, and let

$$K = \max\left\{t_0, v \log \frac{2mn}{\epsilon}\right\}.$$

Then

$$\mathbf{P}\left(\|X\| \geq 2\sigma + \frac{\epsilon v}{\sqrt{mn}} + t\right) \leq (m \vee n) \exp\left(-\frac{t^2}{C_1 (2K)^2}\right) + \epsilon,$$

where σ and C_1 are the same as in Lemma 2.

To apply this result, we need a subexponential tail bound for the Poisson distribution.

Lemma 4. *Let $X \sim \mathrm{Poisson}(\lambda)$. Then*

$$\mathbf{P}(X - \lambda \geq t) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right).$$

For $t \geq \lambda$,

$$\mathbf{P}(X - \lambda \geq t) \leq e^{-3t/8}.$$

The first inequality can be established by approximating the Poisson distribution with mean λ as the sum of k Bernoulli random variables with mean λ/k, applying Bernstein's inequality, and taking $k \to \infty$. The idea for this argument was suggested by an exercise in [14].

Going back to our original problem, we need to bound the operator norm of $Z = \mathcal{A}_\Omega^*(X) - pM$. Note that since we are using a Bernoulli sampling model, the entries of Z are independent. Let $\lambda_{\max} = \max_{i,j} M_{ij}$. Note that for every (i, j), $\mathbf{E}\, Z_{ij} = 0$, and it is easily verified from Lemma 4 that for $t \geq 2\lambda_{\max}$, $\mathbf{P}(|Z_{ij}| \geq t) \leq e^{-t/8}$.

To calculate σ from Lemma 2 in terms of p and M, we note that

$$\begin{aligned}\mathrm{var}(Z_{ij}) &= p\,\mathrm{var}(X_{ij}) + p(1-p)(\mathbf{E}\, X_{ij})^2 \\ &= pM_{ij} + p(1-p)M_{ij}^2,\end{aligned}$$

so we can calculate $\sigma = \sqrt{p}\widetilde{\sigma}$.

B. Minimax lower bounds

In this section, we present results which show that the rate in (3) is optimal (within a multiplicative constant) in the sense of minimax risk. $\widetilde{\sigma}$ is partially determined by the maximal row and column sums of the rate matrix M, which we can think of as the maximal variance of any row or column (without sampling a subset of the entries). Our first lower bound shows that we cannot improve on this term.

Again, complete proofs will appear in a future publication.

Theorem 2. *Let r, k, and ℓ, be positive integers, and take $m = rk$, $n = r\ell$. Let $\lambda_{max} \geq 1/8\ell p$, set $\sigma_1^2 = k\lambda_{max}$, and let*

$$S_1 = \left\{M \in [0, \lambda_{max}]^{m \times n} : \mathrm{rank}(M) \leq r, \right.$$
$$\left. \sqrt{\max_i \sum_j M_{ij}} + \sqrt{\max_j \sum_i M_{ij}} \leq 2\sigma_1\right\}.$$

Then, under a Bernoulli sampling model with sampling probability p,

$$\inf_{\widehat{M}} \sup_{M \in S_1} \mathbf{P}_M\left(\|\widehat{M} - M\|_F \geq \frac{\sqrt{r}\sigma_1}{8\sqrt{2p}}\right) \geq \frac{1}{2} - \frac{8 \log 2}{m \vee n}.$$

The proof is a standard argument by Fano's method (see, e.g., [15], [16]), using block-diagonal matrices whose rows each take one of two values.

This first theorem relies on the fact that the observations are (conditionally) Poisson; the next result, which provides the second part of a matching lower bound to (3), does not depend on the conditional distribution of the observations, and instead shows a fundamental limit in inferring missing matrix entries.

Theorem 3. *Take again $m = rk$, $n = r\ell$. Set $\sigma_2^2 = k\lambda_{max}^2$. Let*

$$S_2 = \left\{M \in [0, \lambda_{max}]^{m \times n} : \mathrm{rank}(M) \leq r, \right.$$
$$\left. \sqrt{\max_i \sum_j M_{ij}^2} + \sqrt{\max_j \sum_i M_{ij}^2} \leq 2\sigma_2\right\}.$$

Suppose $p \geq \frac{1}{2(k \wedge \ell)} = \frac{r}{2(m \wedge n)}$. Then, under a Bernoulli sampling model with probability p (with any conditional distribution on the observations),

$$\begin{aligned}\inf_{\widehat{M}} \sup_{M \in S_2} \mathbf{E}\|\widehat{M} - M\|_F^2 &\geq \frac{r\sigma_2^2}{8} \max\left\{\frac{1}{2}\left\lfloor\frac{1}{2p}\right\rfloor, 1 - p\right\} \\ &\geq \frac{1}{64}\frac{1-p}{p} r\sigma_2^2.\end{aligned}$$

The proof is by Assouad's method (again, see, e.g., [15] or [16]), using a "hypercube" of matrices similar to that in the proof of Theorem 2. We could also use a more complicated argument by Fano's method to get another high-probability lower bound on error (with a somewhat worse constant).

C. When do the upper and lower bounds match?

Within multiplicative constants, the lower bounds of Theorems 2 and 3 match the approximate upper bound of (3). To determine whether this proves optimality, we must consider when the approximation in (3) is accurate.

If we look at the classes of matrices in Theorems 2 and 3, it can easily be shown that the term $\sqrt{p}\widetilde{\sigma}$ dominates (4) if

$$p \gtrsim \frac{r}{m} \max\left\{\log m, \frac{\log^3 m}{\lambda_{\max}^2}\right\}$$

If $\lambda_{\max} \gtrsim \log m$, this reduced to the condition $p \gtrsim \frac{r \log m}{m}$, which is a common assumption in the matrix completion literature.

III. Conclusion and future work

In this paper, we have derived an upper bound in Frobenius norm error for an estimator for Poisson matrix completion, and we have derived a minimax lower bound that matches this upper bound (within a universal constant) for many classes of nonnegative rate matrices. The estimator we use is computationally tractable, and requires significantly fewer assumptions on the underlying matrix than previous results in the literature. Significantly, we impose no lower bounds on the entries of the underlying matrix. This is crucial in many applications (such as topic modelling) where zero or very small Poisson means can be relatively common.

Because we have found upper and lower error bounds in Frobenius norm, the only theoretical improvement remaining for this model and error metric is to try to relax the conditions under which the bounds match (although, as we have seen, they are not too restrictive now). This could potentially come about by reducing the logarithmic term in (4) and/or by finding a logarithmic term to add to the minimax lower bounds.

It would also be interesting to extend the results presented here to matrices that are not exactly low-rank, but are instead "approximately low-rank;" for example, we could consider matrices which are contained in Schatten balls (which, for $q \in [0, 1]$, are sets of matrices for which $\sum_i \sigma_i^q \leq R$, where $\{\sigma_i\}$ is the set of singular values). As mentioned previously, Cao and Xie [6] used the Schatten 1-norm ($q = 1$, or nuclear norm ball); Negahban and Wainwright [1] also examined these classes of matrices.

Another avenue of research would be to examine structured Poisson rate estimation under different, more statistically motivated error metrics. Maximum likelihood methods seem more suitable here than least-squares, but analysis of maximum likelihood estimators has proved difficult for the reasons outlined in Section I-B. It is not clear what kind of structure would be relevant in a different error metric. Low-rank structure seems to work well with a least-squares error framework, but there is *a priori* not much reason to think that it would work similarly well for another metric; for example, the Bhattacharyya distance, for Poisson distributions, is equivalent to (squared) ℓ_2 distance between the *square roots* of the rates. However, the element-wise square root of a low-rank matrix is not, in general, low rank. Thus, this approach may not immediately bear much fruit. However, an analysis of Poisson matrix estimation under alternative error metrics remains an important area for future research.

Acknowledgments

This work was supported, in part, by NSF grant CCF-1350616 and a gift from the Alfred P. Sloan Foundation.

References

[1] S. Negahban and M. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *J. Mach. Learn. Res.*, vol. 13, pp. 1665–1697, May 2012.

[2] V. Koltchinskii, K. Lounici, and A. Tsybakov, "Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion," *Ann. Stat.*, vol. 39, no. 5, pp. 2302–2329, Oct. 2011.

[3] E. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[4] X. Zhou, C. Yang, H. Zhao, and W. Yu, "Low-rank modeling and its applications in image analysis," *ACM Comput. Surv.*, vol. 47, no. 2, 36:1–36:33, Dec. 2014.

[5] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. ACM SIGIR Conf. (SIGIR)*, Berkeley, CA, Aug. 1999.

[6] Y. Cao and Y. Xie, "Poisson matrix recovery and completion," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1609–1620, Mar. 2016.

[7] A. Soni, S. Jain, J. Haupt, and S. Gonella, "Noisy matrix completion under sparse factor models," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3636–3661, Jun. 2016.

[8] A. Soni and J. Haupt, "Estimation error guarantees for Poisson denoising with sparse and structured dictionary models," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Honolulu, HI, Jun. 2014.

[9] J. Lafond, "Low rank matrix completion with exponential family noise," in *Proc. Conf. Learn. Theory (COLT)*, Paris, France, Jul. 2015.

[10] S. Gunasekar, P. Ravikumar, and J. Ghosh, "Exponential family matrix completion under structural constraints," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Beijing, China, Jun. 2014.

[11] G. Raskutti, M. Wainwright, and B. Yu, "Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6976–6994, Oct. 2011.

[12] A. Bandeira and R. van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *Ann. Probab.*, vol. 44, no. 4, pp. 2479–2506, Jul. 2016.

[13] R. Latała, R. van Handel, and P. Youssef, "The dimension-free structure of nonhomogeneous random matrices," *Invent. Math.*, vol. 214, pp. 1031–1080, 3 Dec. 2018.

[14] D. Pollard, "MiniEmpirical," Available at http://www.stat.yale.edu/~pollard/, 2016.

[15] C. Huber, "Lower bounds for function estimation," in *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. L. Yang, Eds., New York: Springer, 1997, ch. 15, pp. 245–258.

[16] B. Yu, “Assouad, Fano, and Le Cam,” in *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. L. Yang, Eds., New York: Springer, 1997, ch. 29, pp. 423–435.